Design and Analysis of Hybrid Recommender System using Machine Learning Techniques

SHORT SYNOPSIS

Submitted in partial fulfillment of the requirement of the degree of

DOCTOR OF PHILOSOPHY

To

Manav Rachna International Institute of Research and Studies
(Deemed to be University)

by

VINEET SHRIVASTAVA
18/Ph.D./069

Under the supervision of
Dr. Suresh Kumar
Professor, CSE

FACULTY OF ENGINEERING & TECHNOLOGY

MANAV RACHNA INTERNATIONAL INSTITUTE OF RESEARCH & STUDIES

SECTOR-43, SURAJKUND –DELHI ROAD, FARIDABAD – 121001

FACULTY OF ENGINEERING & TECHNOLOGY

Certificate

I hereby declare that the short synopsis "**Design and Analysis of Hybrid Recommender System using Machine Learning Techniques.**" being submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy in Computer Science & Engineering under faculty of Engineering and Technology of Manav Rachna International Institute of Research & Studies, Faridabad during the academic year 2020. It is a bona fide record of original work carried out under the supervision of Dr. Suresh Kumar, Professor, Department of Computer Science and Engineering and has not been submitted in full or part to any other university or institute for degree or diploma. Further, it is certified that short synopsis is plagiarism free and proof has been submitted.

VINEET SHRIVASTAVA
18/PhD/069

Signature of the Supervisor
Dr. Suresh Kumar
Professor, CSE
Faculty of Engineering & Technology

(Chairperson, DRC)

Chairman, CDP

ABSTRACT

The recommender system considers a data filtering system that supports user interaction with items and gives preferences to them. The advancement of machine learning and temporal inconsistency among the users leads to competition among the various conventional algorithms resulting in a rapid change of process.

Organizations like Amazon, Netflix, LinkedIn, and Pandora affect the recommender system frameworks to enable shoppers to detect new and closely connected things (items, recordings, occupations, and music), making an excellent customer experience while driving regular income.

The applications on web pages also have the capability of accessing details related to users and products. These details include demographic information of users like time, location, etc. at the time of interactions with the product. They also include product-specific information like the description of the product. These details are analyzed by the recommendation system, as these are actual values. This analysis can lead to finding the user-item matching, and the suggested recommendations.

Recommender systems are mainly categorized into Collaborative, Content-based, and hybrid filtering systems. Collaborative filtering analyzes the historical interaction of users and items. These include the purchase history of the items by the user. Content-based filtering explains the user profile build by the previous likes and ratings of products by a user. The various challenges like scalability, diversity, differential services, and cold start problems arise as a gap for these methods. The hybrid technology combines the benefits of both collaborative and content-based filtering to overcome these challenges.

In this work, the focus is to design and develop an improvement in a hybrid recommender system that can overcome various challenges of Content-based and Collaborative Filtering based methods. These Challenges include cold start problems, scalability of the system, differentiability problems etc. In this research work the power of Machine learning techniques will be utilized in the design of hybrid recommender system. The proposed recommender system will help in recommending multiple items to the users more efficiently as compared to previous algorithms.

Key words: Recommender System, Machine learning, Hybrid systems.

CONTENTS

S. No.	Description	Page No.
1	Introduction	1-2
2	Literature Review	3-5
3	Description of Broad Area 3.1 Characteristics of Machine learning 3.2 Research Gaps 3.3 Research Problem	6-8
4	Objectives of the Study	9
5	Methodology to be Adopted	10
6	Expected Outcome of the Research	11
7	References	12-13

1. Introduction

Recommender systems are defined as a software tool that suggests the items and their features to the users. Earlier, the recommendations were made by the opinions of domain experts having vast amount of knowledge about the subject. As a result of immense growth in digital data in previous decades, the work of recommendation became stressful by using conventional knowledge base techniques. This enormous data lead to the development of software systems that recommend items precisely(Çano and Morisio, 2017).

The Recommender systems emerged as a necessary tool to assist users in accessing services from large sets of merchandise. Netflix provides movie recommendations, Amazon provides product recommendation services, Yahoo! recommends news, and Google gave recommendations of advertisements. Corporate organizations provide net services and recommendation ability together as functionality in the market(Song, Tekin and Van Der Schaar, 2016).

Machine Learning is considered as a sub domain of Artificial Intelligence that deals with the training skills of machine. The Algorithms of machine learning that typically used in the recommender systems are classified into two different categories namely Collaborative filtering and Content based filtering. The Collaborative method works on the interaction between the users and items. It mainly has an idea that people who agreed on previous interaction with the item are again going to like them in the future. The task of algorithms is to search out operation that predicts the utility of that past interaction. The various content-based strategies supports the similarities possess by things throughout their interactions(Isinkaye, Folajimi, & Ojokoh, 2015) .

There are various problems mentioned by the researchers in using the recommendation system strategies. The common issues mentioned are the Cold Start problem, Data sparsity problems, Accuracy problems, Diversity, and Scalability. Various researches were conducted with the help of machine learning to reduce the effectiveness of these challenges. These techniques still have a scope of extensions and futuristic measures that used to develop a better recommendation system(Song, Tekin and Van Der Schaar, 2016).

The author in this research compares various algorithms that previously used to overcome the mentioned problems and find out the various changes by comparing more contexts of data. The system investigates the machine learning algorithm that is best fitted to a given data. For better evaluation of large–scale contextual data, a hybrid approach that consists of various

machine learning methods is proposed to provide more accuracy than the previously designed system.

2. Literature Review

The core operation of any recommender system is to predict an item that may be found helpful by a user. The basis for the recommendation, however, varies from one recommender system to another. There exist several taxonomies to distinguish between different approaches to make a recommender system. Previously, the recommender system follows three basic approaches based on characteristic information and User item interactions. These techniques are Content-based filtering, Collaborative based filtering, and Hybrid filtering.

Asim A, Skander E. & R. kohli (2000) developed a Hierarchical Bayesian Recommendation system. The products for a recommendation in this system were new release and old release movies. The system, with the help of statistical methods based on customer ratings on products, was developed. Customized recommendations to the customers given with the help of a regression-based approach. The results show a better recommendation for the movie having higher ratings. The authors mentioned that for the recommendation of lower rating movies, this model was less desirable.

Gediminas Adomavicius, Bamshad Mobasher, Francesco Ricci & Alex Tuzhilin (2010) have given the concept of context in the recommender system. In their work, the authors mentioned the importance of contextual factors and change in them according to time. The authors also suggested the limitations of this system. The problem described in these systems was identifying which context is to be considered is still an area for studies. They also mentioned that the classification is only an initial and general high-level structure for characterizing the multi-faceted subject of contextual knowledge in recommendation systems.

Linqi Song, Cem Tekin & Mihaela van der Schaar (2013) developed a recommender system algorithm based on the recommendation at the item cluster level. The authors mentioned various challenges like scalability, diversity, differential services, and cold start problems. This recommendation system worked on an associate algorithmic program that considers item preferences of users, which uses a square measure technique that supported their interaction with them. The authors in their system constructed an online item cluster tree and made a recommendation at the item-cluster level. They also introduced Adaptive Clustering Recommendation algorithm (ACR) for improving learning speed. This ACR technique is used to select adaptive cluster size giving balance to the recommendation and reducing cluster size. A comparison with another context-free algorithm has been made in this work. The limitations of this system occurred when this algorithm is used in a distributed manner.

Yao Lu, Sandy El Heou, Denis Gillet (2013) gave a hybrid recommender system for job seekers. They used a Content-based Similarity Computing and Ranking algorithm for the construction of the system. They used candidate, employer, and job fields and computed profile-based similarity and by using Personalized Multi-Relational Page Rank and indicated the importance of a particular page. The authors gave a better recommendation of this approach than previous approaches.

Negar Hariri, Bamshed Mobasher & Robin Burke (2014) have developed a system that adapts the contextual changes in interactive recommendations. The authors used the Multi-Arm Bandit algorithm and extended them to study the variation of the user's feedback due to contextual changes. If the change is noted, the algorithm based on Thompson sampling priorities the result. The algorithms prioritize the current context and depend on the user's feedback.

Maria R. Lee, Tsung Teng Chen & Ying Shun Chai (2016) have developed a Recommender System based on hybrid techniques using data from social networking sites for recommending Taiwan premiere movies. They have used Fb (API) to extract data from social media fan pages. The data preprocessing stage involves a collaborative filtering technique and applied multinomial logistic regression. They have evaluated results in two parts i.e., Cross-Validation and Online system questionnaire evaluation process. The accuracy is calculated for various genres, and the perception of users is rated by using the questionnaire evaluation process based online. This Recommender System is analyzed text only in traditional Chinese language, and experiments on other languages are not performed. Sufficient data on fan pages is required for the recommender system to work appropriately.

Agata Nawroka, Andrzej Kot & Marcin Nawrocki (2018) have studied various issues related to the application of Machine Learning in recommender systems. The study was conducted on data present in the Movie lens portal. The authors created algorithms that calculate prediction ratings. This algorithm was based on similarity if users and the objects. The errors that were considered in the prediction are RMSE and MAE, and the results show that these errors were less. The author mentioned that the system is effective only when a certain number of minimum ratings by users are considered.

Geetha G., Safa M., Fancy C., Saranya D. (2018) have developed a movie recommender system that was based on hybrid filtering for précised recommendation. The authors centered on the mix of the benefits of Content-based and collaborative based mostly filtering ways. The system allows the users to choose from the given set of attributes and then on the basis of the cumulative weight of attributes, and also, by using the k – mean algorithm, the system

was able to recommend a list of movies. The system used only a small set of users that are based on an informal evaluation.

The work in the area of recommendation system based on the techniques of the Content base, Collaborative, and hybrid technologies. The algorithms used by authors commonly focused on the development of methods of more accuracy. More studies can achieve further exploration regarding overcoming the disadvantages like Cold – Start Problems, Scalability, sparsity problems, etc. These literature works give the base for experimentation with the help of Machine learning algorithms for achieving higher accuracy in recommender systems.

3. DESCRIPTION OF BROAD AREA

Machine Learning is considered as a sub-domain of AI that deals with computation for extracting new information from the prevailing one. Most of the work previously accomplished by personalities were with success replaced by a Machine with the flexibility of fewer errors. Machine learning is influencing the ability of knowledge and self-capability in machines.

3.1 Characteristics of Machine Learning

The following characteristics describe Machine Learning.

3.1.1. The capacity to perform automated Data Visualization.

Every day in various activities, a tremendous amount of data is increasing in multiple fields regularly. These fields may include business data or data generated by ordinary people. For determining the relationship among the data, the visualizing perspective not only helps to make better decisions but also helps in building confidence. Machine learning tools help the data to be visualized in a form that the relationship between the data is understandable. This technique could lead the businesses to grow by getting clear insights into the data details (Li, Fang, Mukhopadhyay, Saykin & Shen,2018).

3.1.2. The ability for Automation of repetitive tasks

Repetitive tasks can define the same functions that are performed regularly in an organization. By automation of these tasks, the time and efforts of manual work is reduced by a large extent. Machine learning techniques help to automate these repetitive tasks. A repetitive job like accounting in financial sectors requires time, and by automation, the time saved. This automation applied chatbots by making predictions with the help of machine learning algorithms. These are also helpful in the management of the expenditure and simplifying invoice facility to the customers (Ruvini & Dony 2000).

3.1.3. Machine Learning on the Internet of things (IoT)

Nowadays, the companies are manufacturing products that are IoT enabled. For maximizing benefits from the machine learning techniques would be more convenient as these techniques will boost up the Accuracy of the IoT enabled products. Also, machine learning techniques will help the organizations to determine productivity of the device. In the future, a

combination of IoT and machine learning will boost strategic business decision making efficiency to a great extent (Moh & Raju,2018).

3.1.4. Increased Accuracy in data analysis

In conventional approaches, data analysis based on the trial and error methods were used. This approach was not feasible when working over a vast number of heterogeneous datasets. Machine learning algorithms provide useful measurements for this vast number of datasets. Hence with the help of data-driven models and various algorithms, it can generate increased accuracy results for the system (Ivanytska, Zubyk, Ivanov & Domracheva,2019).

.

3.1.5. Business intelligence at its best

The vast amount of analytical work, when combined with machine learning techniques, provides creative, strategic results in the decision making of business in industries. Today, most of the sectors, whether its financial services or medical services the machine learning evolves as an effective technique. Various algorithms predicted the data by using models. Machines are using learning techniques for predicting accuracy of multiple datasets. These algorithms work differently with a common goal of making predictions.

One of the major application areas of Machine learning is Recommender System. In the relation of the Recommender system approaches the various techniques like Collaborative Filtering and Content-based filtering can be categorized differently. e.g., Collaborative Filtering based on probabilistic and non-probabilistic algorithms and Content-based Filtering based on Machine learning. The various drawbacks like Cold start, Sparsity, etc. have occurred in the conventional approaches. Hence by proposing a Hybrid method that combines multiple Machine learning algorithms can be useful to reduce these drawbacks. The various organizations have started the combination and implemented them on their products. News Dude uses a combination of Naïve Bayes and K nearest neighbor algorithms for their recommendations so that these Machine learning algorithms would play a significant role in the accurate predictions made by the Recommender Systems (Valter, Lindgren & Prasad, 2017).

3.2 Research Gaps

It has been observed that the issues like System bias, Myopic Recommendations, Cold-Start problems, Differentiability, Scalability are yet to be explored widely. These areas by the means of recommendation methods and applying machine learning techniques can help give

more accurate results. Furthermore, the various authors suggested extending the previously proposed solutions for better evaluation. The work can be explored by experimenting on hybrid recommender systems so that more contexts can be included and more personalized recommendations can be provided. Currently, the systems can be categorized according to the following problems like system bias and myopic recommendations (Ge & Persia, 2017).

3.2.1. System Bias

Only observe feedback on items recommended by current systems, but not others.

3.2.2. Myopic Recommendation

Recommendations that are made to the users are likely to lead to immediate response instead of long-term user utility.

3.3 Research Problem

These Research gaps along with challenges like Cold–Start problem, differentiability, diversity, Scalability are always considered in the work areas of the Recommender system. Hence, they still have to be studied more and work can be done in the direction to overcome these challenges. Hence by combining various recommendation methods and developing a hybrid of them by using machine learning algorithms are the perspective of this research work.

4. Objectives of the Study

- To Understand the concept of recommendation system.
- To Compare the performance of existing recommender systems.
- To Propose a hybrid Recommender system that is more accurate and efficient in terms of recommendations.
- To Design an efficient hybrid Recommender System based on machine learning techniques.
- To Implement and analyze the results of the proposed Hybrid Recommender system with Various Datasets.

5. Methodology

The methodology to be adapted is divided into multiple phases as below:

Phase 1: Literature Review
In this phase, the literature related to design and development of recommendation systems using various machine learning techniques will be reviewed.

Phase 2: Analysis of existing recommender systems
In this phase, the existing recommender systems will be analyzed. The various challenges in conventional recommender systems will be identified.

Phase 3: Design of Hybrid Recommender system
Based on the challenges identified in phase 2, design of a hybrid recommender system will be proposed, so that the challenges of existing recommender systems can be dealt more efficiently.

Phase 4: Evaluation of the proposed Efficient Hybrid Recommender system
The performance of the proposed efficient Hybrid Recommender system will be evaluated with experiments on various datasets. Finally, the efficiency of the proposed system will be compared with the existing recommender systems.

6. EXPECTED OUTCOME OF THE RESEARCH

In this research work, an efficient hybrid recommender system will be designed and developed that will mitigate the challenges like scalability, diversity, differential services, and cold start problems more effectively.

REFERENCES

Adomavicius, G., & Tuzhilin, A. (2010). Context-Aware Recommender Systems. *Recommender Systems Handbook*, 217–253. doi: 10.1007/978-0-387-85820-3_7

Ansari, A., Essegaier, S., & Kohli, R. (2000). Internet Recommendation Systems. *Journal of Marketing Research*, *37*(3), 363–375. doi: 10.1509/jmkr.37.3.363.18779.

Çano, E., & Morisio, M. (2017). Hybrid recommender systems: A systematic literature review. *Intelligent Data Analysis*, 21(6), 1487–1524. doi: 10.3233/ida-163209

Geetha, G., Safa, M., Fancy, C., & Saranya, D. (2018). A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System. *Journal of Physics: Conference Series, 1000*, 012101. doi: 10.1088/1742-6596/1000/1/012101

Ge, M., & Persia, F. (2017). A Survey of Multimedia Recommender Systems: Challenges and Opportunities. *International Journal Of Semantic Computing*, *11*(03), 411-428. doi: 10.1142/s1793351x17500039

Hariri, N., Mobasher, B., & Burke, R. (2014). Context adaptation in interactive recommender systems. *Proceedings of the 8th ACM Conference on Recommender Systems - RecSys* 14. 41-48.doi: 10.1145/2645710.2645753

H. Li, S. Fang, S. Mukhopadhyay, A. J. Saykin and L. Shen (2018), "Interactive Machine Learning by Visualization: A Small Data Solution," IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, 3513-3521, doi: 10.1109/BigData.2018.8621952.

Isinkaye, F., Folajimi, Y., & Ojokoh, B. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3),261–273.doi: 10.1016/j.eij.2015.06.005

Ivanytska, A., Zubyk, L. Ivanov, D. & Domracheva,K.(2019). "Study of Methods of Complex Data Analysis that Based on Machine Learning Technologies," *IEEE International Conference on Advanced Trends in Information Theory (ATIT)*, Kyiv, Ukrainep, 332-335, doi: 10.1109/ATIT49449.2019.9030531

Lee, M. R., Chen, T. T., & Cai, Y. S. (2016). Amalgamating Social Media Data and Movie Recommendation. *Knowledge Management and Acquisition for Intelligent Systems Lecture Notes in Computer Science*, 141–152. doi: 10.1007/978-3-319-42706-5_11

Lu, Y., Helou, S. E., & Gillet, D. (2013). A recommender system for job seeking and recruiting website. *Proceedings of the 22nd International Conference on World Wide Web - WWW 13 Companion*. 963–966.doi: 10.1145/2487788.2488092

Moh, M. & Raju, R. (2018). Machine Learning Techniques for Security of Internet of Things (IoT) and Fog Computing Systems. International *Conference on High Performance Computing & Simulation (HPCS)*, Orleans 709-715, doi: 10.1109/HPCS.2018.00116.

Nawrocka, A., Kot, A., & Nawrocki, M. (2018). Application of machine learning in recommendation systems. *19th International Carpathian Control Conference (ICCC)*. 328-331.doi: 10.1109/carpathiancc.2018.8399650

Ruvini, Jean-David & Dony, Christophe. (2000). APE: Learning User's Habits to Automate Repetitive Tasks. *5th international conference on Intelligent user interfaces* 229-232. doi:10.1145/325737.325854.

Song, L., Tekin, C., & Schaar, M. V. D. (2016). Online Learning in Large-Scale Contextual Recommender Systems. *IEEE Transactions on Services Computing*, 9(3), 433–445. doi: 10.1109/tsc.2014.2365795

Valter, P., Lindgren, P. & Prasad, R. (2017), "Artificial intelligence and deep learning in a world of humans and persuasive business models," *Global Wireless Summit (GWS)*, Cape Town, 2017, 209-214, doi: 10.1109/GWS.2017.8300487.

MANAV RACHNA
Vidyanatariksha

Dr. O.P. Bhalla Central Library
Manav Rachna International Institute of Research and Studies
(A deemed-to-be-university under section 3 of UGC Act 1956)

SIMILARITY VERIFICATION CERTIFICATE

Title: Design and Analysis of a Hybrid Recommender System using Machine Learning techniques

Researcher: Vineet Shrivastava Registration Number: 18/PhD/069

Supervisor: Dr. Suresh Kumar Department: CSE

This is to certify that the above mentioned Short Synopsis/~~Thesis~~ was scanned for similarity index thro. Turnitin Software on 30/Jan/2020 withsubmission ID 1248544213 and the similarity index was found to be 4% %.The complete report running into 18 Pages, 3040 Word count with File Size 1570 KBis being forwarded to the Research Supervisor/DRC/DRB for review.

Date: 3/2/2020

Dr. Vikram Kr. Sharma
University Librarian
Dr. O.P. Bhalla Central Library
Manav Rachna International
Institute of Research & Studies
Faridabad
University Librarian

The complete report of the above Short Synopsis/~~Thesis~~ has been reviewed and the similarity index is below the accepted norms as per the University Grants Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions) Regulations 2018. The Short Synopsis/~~Thesis~~ may be considered for further processing

Vineet Shrivastava 3/2/2020
Research Scholar

Research Supervisor
03/02/2020

Scanned By Scanner Go

Short Synopsis VINEET SHRIVASTAVA Jan.30,2020

ORIGINALITY REPORT

4%	2%	2%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University College London Student Paper	1%
2	Submitted to Royal Melbourne Institute of Technology Student Paper	<1%
3	badrit.com Internet Source	<1%
4	Submitted to University of Hertfordshire Student Paper	<1%
5	Submitted to Guru Nanak Dev Engineering College Student Paper	<1%
6	link.springer.com Internet Source	<1%
7	mycomputer.computer.org Internet Source	<1%
8	journalogy.net Internet Source	<1%

Dr. Vikram Kr. Sharma
Dr. O P. Bhalla Central Library
Institute of Research & Studies
Faridabad

Scanned By Scanner Go

9	[redacted].science.gov Internet Source	<1%
10	Submitted to Multimedia University Student Paper	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		

Scanned By Scanner Go